REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
	Technical Report	-

4. TITLE AND SUBTITLE
Stein's Phenomenon and Nanoparticle Characterization

5a. CONTRACT NUMBER
W911NF-12-1-0422

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Richard Charnigo, Limin Feng, Mathieu Francoeur, Cidambi Srinivasan

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of Kentucky
University of Kentucky Research Foundation
500 South Limestone Street
Lexington, KY 40526 -0001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
61653-MA-II.15

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
Motivated by the possibility of engineering nanostructured surfaces for efficient nano-thermophotovoltaic power generation, we investigate whether and how Stein's phenomenon may effectively be incorporated into nanoparticle characterization based on scattering data. The statistical innovation in our approach is to employ a form of nonlinear shrinkage in the compound estimation of scattering profiles and their derivatives, such that the estimation error is reduced without destroying the self-consistency property of compound estimation (i.e., that the estimated

15. SUBJECT TERMS
compound estimation, local regression, nano-thermophotovoltaic power generation, nonparametric regression, self-consistency, Stein's phenomenon

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Richard Charnigo
UU	UU	UU	UU		19b. TELEPHONE NUMBER: 859-218-2072

Report Title

Stein's Phenomenon and Nanoparticle Characterization

ABSTRACT

Motivated by the possibility of engineering nanostructured surfaces for efficient nano-thermophotovoltaic power generation, we investigate whether and how Stein's phenomenon may effectively be incorporated into nanoparticle characterization based on scattering data. The statistical innovation in our approach is to employ a form of nonlinear shrinkage in the compound estimation of scattering profiles and their derivatives, such that the estimation error is reduced without destroying the self-consistency property of compound estimation (i.e., that the estimated derivatives are precisely the derivatives of the estimated scattering profiles). Our numerical experiments found that the estimation error may be reduced by up to 7%, 25%, and 31% respectively in estimating the scattering profiles, their first derivatives, and their second derivatives. Surprisingly, the reduced estimation error does not translate into demonstrably superior correct classification rates for nanoparticle configurations. Speculations on the findings and promising avenues for future research are discussed.

Technical Report:

Stein's Phenomenon and Nanoparticle Characterization

Richard Charnigo[1,*], Limin Feng[1], Mathieu Francoeur[2], and Cidambi Srinivasan[1]

[1] Department of Statistics, 725 Rose Street,

University of Kentucky, Lexington KY USA 40536

E-Mail: RJCharn2@aol.com, limin.feng@uky.edu, srini@ms.uky.edu

[2] Department of Mechanical Engineering, 50 S. Central Campus Dr.,

University of Utah, Salt Lake City UT 84112

E-Mail: mfrancoeur@mech.utah.edu

*Corresponding author

Abstract

Motivated by the possibility of engineering nanostructured surfaces for efficient nano-thermophotovoltaic power generation, we investigate whether and how Stein's phenomenon may effectively be incorporated into nanoparticle characterization based on scattering data. The statistical innovation in our approach is to employ a form of nonlinear shrinkage in the

compound estimation of scattering profiles and their derivatives, such that the estimation error is reduced without destroying the self-consistency property of compound estimation (i.e., that the estimated derivatives are precisely the derivatives of the estimated scattering profiles). Our numerical experiments found that the estimation error may be reduced by up to 7%, 25%, and 31% respectively in estimating the scattering profiles, their first derivatives, and their second derivatives. Surprisingly, the reduced estimation error does not translate into demonstrably superior correct classification rates for nanoparticle configurations. Speculations on the findings and promising avenues for future research are discussed.

Key words: compound estimation, local regression, nano-thermophotovoltaic power generation, nonparametric regression, self-consistency, Stein's phenomenon

1 Introduction

Motivation for nanoparticle characterization. The current world energy consumption is approximately 14 TW, among which less than 1% is coming from clean and renewable sources (Baxter et al 2009). By 2050, this global demand is expected to reach 25 to 30 TW. In order to minimize the environmental impacts of this increasing energy consumption, experts estimate that about 20 TW should come from carbon-free renewable energy resources. Baxter et al (2009) pointed out the importance of nanoengineering to develop low-cost and high-efficiency renewable energy technologies, and emphasized solar thermophotovoltaic (TPV) power generators that could greatly benefit from nanoscale design. In such devices, solar irradiation is absorbed by a radiator, which, in turn, re-emits thermal radiation in a spectrally selective fashion toward a cell generating electricity. TPV power generators are not restricted to solar

application, as any kind of heat source can be used. For instance, direct thermal-to-electrical energy conversion via TPV devices could diminish waste heat (58% of the 110 EJ consumed annually in the US is lost to heat, Bergman et al 2011) in various systems such as combustion chambers, photovoltaic cells and personal computers.

TPV power generation is limited by Plancks blackbody distribution. To improve TPV performances, Whale and Cravalho suggested spacing the radiator and the cells by a sub-wavelength vacuum gap (Whale 1997, Whale and Cravalho 2002). At sub-wavelength distances, radiation heat transfer is in the near-field regime, such that energy exchange can exceed by a few orders of magnitude the blackbody predictions. The literature on nano-TPV suggests that power generation can be enhanced by a factor of 20 to 30 when the gap between the radiator and the cell is a few tens of nanometers (Laroche et al, 2006, Park et al 2008). This conclusion, however, was made by ignoring the thermal impacts on nano-TPV performances. Using a multi-physics model combining near-field thermal radiation, heat conduction and charge transport in TPV cells, Francoeur et al (2011) showed that that naively spacing the radiator and the cell by a few tens of nanometers is not a viable solution. This is because the near-field radiative heat transfer enhancement is broadband, contributing not only to increased photocurrent generation but also to heating of the cell due to absorption by the lattice and the free carriers, electron-hole pair recombination and thermalization thus leading to low nano-TPV performances.

The key to developing viable, highly efficient nano-TPV devices is to tune near-field radiation heat transfer from the radiator to the cell so that most of the radiative energy is used for generating electricity. In practice, lower-temperature nano-TPV devices ($\sim$ 320 - 500 K) will likely be easier to implement. Polar crystals such as silicon carbide (SiC) and cubic boron

nitride (cBN), supporting surface phonon polaritons (SPhPs) in the infrared (Narayanaswamy and Chen 2003, Mulet et al 2002, Joulain et al 2005, Francoeur et al 2010), can potentially be employed for the radiator. Indeed, the near-field thermal spectrum emitted by polar crystals is quasi-monochromatic near the resonant frequency of SPhPs. On the other hand, low energy (i.e., long wavelength) SPhPs modes in polar crystals ($\sim$ 0.12 eV) are difficult to convert into electricity due to current TPV cell technological limitations. Instead, quasi-monochromatic near-field thermal sources with resonance in the near-infrared (NIR, $\sim$ 0.5 - 0.8 eV), which are difficult to find in nature (West et al, 2010), are needed for viable, highly efficient nano-TPV power generation. NIR resonance can be obtained using nanoparticles ($\sim$ 5 nm - 200 nm) with high electric permittivity in that spectral band, such as silicon, germanium, and alumina (Wheeler et al 2005, Garcia-Etxarri et al 2011, Petersen et al 2013). Indeed, a large collection of such dielectric particles having strong enough electric and magnetic dipole resonances can induce "effective" negative permittivity and permeability leading to surface polariton resonance. If the nanoparticles are coated on a low-emitting substrate, then the radiative heat flux is expected to be large near resonance but small at other frequencies, since thermal emission at frequencies other than the resonant one is proportional to the volume of the emitter.

Fine tuning of the near-field thermal spectrum can be achieved by varying nanoparticle properties (e.g., size, shape, and agglomeration level); tuning resonance slightly above the absorption bandgap is anticipated to optimize nano-TPV performance. The viability of nano-TPV power generation therefore depends crucially on the ability to control nanoparticle properties. In turn, controlling these properties requires the capacity to characterize nanoparticles non-intrusively and in real time. As the near-field thermal spectrum emitted is very sensitive to surface characteristics, a solid non-intrusive characterization framework is crucial

for manufacturing selective thermal emitters. The application of selective near-field thermal emitters to nano-TPV power generators will allow low temperature waste heat recovery in a variety of electronic devices, such as cell phones and photovoltaic cells. Among the devices that can be powered by nano-TPV systems or whose functionality can be improved are several used by the Army in its mission to provide prompt, sustained land dominance in military operations: computers, communication devices (e.g., radios) and night vision cameras. Worth noting is that the design of structures with user-defined thermal radiative properties will have other applications, such as optical cloaking. In particular, the ability to control the thermal spectrum will pave the way for developing materials invisible to infrared cameras. Such passive infrared camouflage is also relevant to military operations.

A scattering paradigm for nanoparticle characterization. Current methods such as scanning electron microscopy, scattering tunneling microscopy and atomic force microscopy all provide excellent visual imaging of nanoscale materials, yet are unable to do so without having some effect on the sample. They are also unable to give precise characteristics in real time such as size, arrangement and composition. A possible way to perform non-intrusive characterization is to illuminate nanostructures with visible light. The energy and polarization state of light can be described by the four-component Stokes vector (Bohren and Huffman 1998). When radiation interacts with matter, its intensity, polarization state, and direction of propagation are modified. The general relationship between an incident Stokes vector $\mathbf{I}_i$ and a scattered Stokes vector $\mathbf{I}_s$ is $\mathbf{I}_s = [M_{ij}]\mathbf{I}_i$, where $[M_{ij}]$ is the 4 by 4 Mueller scattering matrix (Bohren and Huffman 1998). A Mueller matrix can be associated with any medium modifying the intensity and polarization state of light. Therefore, by measuring the Mueller scattering matrix elements M_{ij} due to nanoparticle scattering, one can infer nanoparticle properties using an inversion

procedure.

Such a non-intrusive characterization technique was proposed by Mengüç and Manickavasagam (1998). Their framework entails illuminating particles with a propagating radiation beam in the visible spectrum, after which far-field scattered light is collected. The scattered intensity is typically insensitive to the nanoparticles properties (Francoeur et al 2007), while the change of polarization after scattering usually is. For that reason, the method entails recovering the Mueller scattering matrix that contains information about the change of polarization (Bohren and Huffman 1998). However, their technique applies to particles larger than a few hundred nanometers, since scattering is too low for smaller particles, while we seek to characterize nanoparticles from 5 to 200 nm. A way to circumvent this problem, while keeping the idea of measuring the change of intensity and polarization state of light via the Mueller scattering matrix elements, is to illuminate nanoparticles deposited on a surface with evanescent waves generated by total internal reflection of an external radiation beam.

The nanoparticle characterization scattering paradigm has been described by Francoeur (2010) and Short (2013). Briefly, nanoparticles to be characterized are deposited on a substrate. Elliptically polarized radiation with wavelength in the visible band (400 to 700 nm) is directed to the interface between the substrate and the air, at which the nanoparticles are located, at an incidence angle exceeding the critical angle for total internal reflection (Novotny and Hecht 2006). The nanoparticles then scatter the evanescent waves generated at the substrate-air interface, and light is again polarized before being recovered at various angles in the far field. The characterization framework has three essential parts: the forward model for predicting evanescent wave scattering by nanoparticles on a surface, the database of scattering profiles and the inversion algorithm. The forward model is used to build the

scattering profile database, a database that correlates varying characteristics to their corresponding scattering profiles. Then, using the database assembled from the forward model, an inversion algorithm is employed to back-calculate characteristics from scattering profiles measured experimentally from an unknown source.

Various methods can be used for predicting scattering of evanescent waves by nanoparticles on a surface. We used a transfer matrix approach for building a database of scattering profiles (Venkata et al 2007, Charnigo et al 2007, Charnigo et al 2011, Charnigo et al 2012). This approach is, however, limited to spherical nanoparticles. An alternative forward model that is being evaluated for the characterization framework is the Discrete Dipole Approximation with Surface Interaction (DDA-SI) (Short 2013, Loke and Menguc 2010, Loke et al 2011, Short et al 2013). The DDA-SI is based on discretizing the objects into cubical sub-volumes behaving as electric point dipoles. Since calculations are made for each individual sub-volume, the DDA-SI can easily handle any complex-shaped inhomogeneous scatterers.

Experimentally, pre- and post-polarization of light is performed using a set of linear polarizer (LP) and quarter-wave plate (QWP) before illuminating the sample and after scattering by the nanoparticles. The relationship between the incident and scattered Stokes vectors can be written as (Francoeur 2010): $\mathbf{I}_s = \{[M_{ij}]_{LP2} \cdot [M_{ij}]_{QWP2} \cdot [M_{ij}] \cdot [M_{ij}]_{QWP1} \cdot [M_{ij}]_{LP1}\}\mathbf{I}_i$, where the subscripts LP1 and QWP1 imply that the optical components are located before illumination of the sample while the subscripts LP2 and QWP2 refer to the components used to polarize light after scattering. The Mueller scattering matrices for LPs and QWPs are known (Bohren and Huffman 1998), while the unknown Mueller matrix due to nanoparticle scattering $[M_{ij}]$ is determined from multiple experiments since only the intensity I (first component of the Stokes vector) is measured. The intensity is a linear combination of N Mueller scattering

matrix elements, where N depends on the equivalent Mueller matrix multiplying $\mathbf{I}_i$ in the above equation. Therefore, to retrieve N Mueller scattering matrix elements from intensity measurements, one must perform N distinct experiments. This system of N equations with N unknowns can be constructed by orienting the LPs and QWPs at different angles (Francoeur 2010). More details about the inversion are provided next.

Using derivatives of scattering profiles. The basic idea of the scattering paradigm, then, is to compare scattering profiles recently acquired from nanoparticles whose properties are unknown to established scattering profiles previously obtained from nanoparticles whose properties are known. From a statistical viewpoint, this is a pattern recognition problem. There are different ways in which scattering profiles may be compared, but one is L^2 distance. In words, the unknown properties are guessed to match the known properties corresponding to the established scattering profiles that are closest, in terms of integrated squared deviation, to the recently acquired scattering profiles.

However, low-frequency high-amplitude content of the scattering profiles (i.e., global trends) contributes more to L^2 distance between scattering profiles than does low-amplitude high-frequency content (i.e., local trends). If low-frequency high-amplitude content varies substantially with the nanoparticle properties, then this is quite desirable. On the other hand, if low-amplitude high-frequency content varies more substantially with the nanoparticle properties, then the aforementioned comparison may not be a good solution to the pattern recognition problem. This observation prompted Charnigo et al (2007) to examine whether looking at *derivatives* of the scattering profiles may lead to a better solution. Of course, the derivatives must be *estimated*, and this is accomplished using a nonparametric regression approach called compound estimation (Charnigo and Srinivasan, 2011a). We briefly review this approach in

the Methods section of the present technical report. Charnigo et al (2007) indeed discovered that, under certain circumstances, examining the distance between *derivatives* of the scattering profiles could yield higher correct classification rates than examining the distance between the scattering profiles themselves. Although they considered L^1 distance (integrated absolute deviation) rather than L^2 distance, their reasoning also applies with L^2 distance.

Of course, one may obtain different classifications with first derivatives than with second derivatives. While choosing a single order of derivative may be appealing for its simplicity, considering multiple orders of derivatives simultaneously may yield higher correct classification rates. For example, one can attempt to minimize a weighted sum of the distance between first derivatives and the distance between second derivatives. Indeed, one can attempt to minimize a weighted sum of the distance between first derivatives, the distance between second derivatives, and the distance between undifferentiated scattering profiles. This idea was explored in detail by Charnigo et al (2011). Subsequently, Charnigo et al (2012) demonstrated how to use derivatives of scattering profiles to obtain confidence intervals (or, to be more technically precise, Bayesian credible intervals) for nanoparticle properties that may reside in a continuum (for instance, diameter, which may reside in a continuum between 5 and 100 nanometers).

Motivating and describing the present work. As noted above, using derivatives of scattering profiles for nanoparticle characterization necessarily entails estimation of these same derivatives. This is actually a high-dimensional estimation problem, for a few reasons: multiple response variables are considered, corresponding to different elements of the Mueller scattering matrix (Cf. Bohren and Huffman, 1998); multiple orders of derivatives are considered; and, numerous observation angles are considered (in fact, strictly speaking, they reside in a continuum). The nonparametric regression approach of compound estimation effectively reduces

the dimensionality inherent to the observation angles by focusing attention on a comparatively smaller number of "centering points", but even then the dimensionality is fairly high. The concept of centering points is explained in the Methods section of this technical report when we describe compound estimation.

In the comparatively simpler problem of estimating the mean of a multivariate normal distribution, Stein (1956) demonstrated that the sample mean was not generally the best estimator available in the sense of minimizing the expected squared error. This counterintuitive result is known as Stein's phenomenon and has a profound connection to the dimensionality of the estimation problem. When there are three or more dimensions, a nonlinear adjustment to the sample mean can reduce the expected squared error, and the reduction becomes more pronounced as the number of dimensions increases. James and Stein (1961) gave an explicit prescription for the nonlinear adjustment, and further variants were subsequently proposed by other authors including Efron and Morris (1973). Charnigo and Srinivasan (2011b) provide a detailed review of Stein's phenomenon. However, the key idea for the present technical report is that Stein's phenomenon has the potential to improve the estimation of derivatives of scattering profiles due to the aforementioned high dimensionality.

Actually, there are several ways in which Stein's phenomenon may be exploited, since as noted above there are several sources of the high dimensionality: multiple response variables, multiple orders of derivatives, and multiple centering points. Herein we consider two of these ways, namely: Steinization simultaneously across multiple response variables and multiple orders of derivatives at each centering point; and, Steinization across multiple centering points for each response variable and order of derivative. We describe these modes of applying Stein's phenomenon in the Methods section and illustrate them via Figures 1 and 2 in the present

technical report.

The hope is that, if the scattering profiles and their derivatives can be estimated more accurately, particularly from data heavily contaminated by disturbances to proper experimental conditions and stochastic noise (Cf. Figure 3), then the resulting characterization of nanoparticles may be more accurate. We investigate whether this is so by conducting numerical experimentation, as described in the Methods section. The findings from the numerical experimentation are summarized in the Results section and Tables 1 through 4. We conclude this technical report with a Discussion of the findings, including implications for ongoing research.

2 Methods

Scattering profiles. Figure 1 of Charnigo et al (2011) illustrates the experimental apparatus from which radiation is scattered for the purpose of characterizing nanoparticles whose properties are unknown. This experimental apparatus is also described in Section 1 of Charnigo et al (2011). Letting $\theta \in [0, 180]$ denote the angle in degrees at which scattered radiation is observed in the far field, we may summarize the relationship between incident and scattered Stokes vectors for a nanoparticle using four functions $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$, and $M_{34}(\theta)$. These functions, which are called scattering profiles, describe the flux and polarization state of the scattered radiation in relation to the observation angle. See Bohren and Huffman (1998) for details regarding underlying physics and Aslan et al (2005), Videen et al (2005), and Venkata et al (2007) for information on numerical computation. In what follows, we let $M(\theta)$ denote any one of $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$, and $M_{34}(\theta)$ generically when no distinction is required.

A laboratory experiment (or numerical simulation thereof) does not actually entail acqui-

sition of $M(\theta)$ but rather samples $M(\theta)$ at a grid of points $\{\theta_1, \ldots, \theta_G\}$, where G denotes the size of the grid. In this technical report, we shall take $G := 177$ and $\theta_j := j + 1$ for $j \in \{1, 2, \ldots, 177\}$. In addition, there may be some systematic and stochastic errors in the sampling of $M(\theta)$, so that, instead of acquiring $\{M(\theta_1), M(\theta_2), \ldots, M(\theta_G)\}$, one acquires $\{M(\theta_1) + \epsilon_1, M(\theta_2) + \epsilon_2, \ldots, M(\theta_G) + \epsilon_G\}$ for some "perturbations" $\epsilon_1, \epsilon_2, \ldots, \epsilon_G$. We emphasize that these perturbations are unknown in practice, so they cannot simply be subtracted out to acquire $\{M(\theta_1), M(\theta_2), \ldots, M(\theta_G)\}$. In other words, $\{M(\theta_1), M(\theta_2), \ldots, M(\theta_G)\}$ are themselves unknown.

Compound estimation. Charnigo and Srinivasan (2011a) developed a nonparametric regression method called compound estimation, useful for estimating a function such as $M(\theta)$ from perturbed data $\{M(\theta_1) + \epsilon_1, M(\theta_2) + \epsilon_2, \ldots, M(\theta_G) + \epsilon_G\}$ acquired on a grid of points. No assumptions are made about the precise mathematical form of $M(\theta)$, but $M(\theta)$ is assumed to be a smooth function of θ. In this technical report, we shall assume that $M(\theta)$ has (at least) three continuous derivatives with respect to θ. Compound estimation differs from other nonparametric regression methods, such as local regression (Cleveland, 1979; Loader, 1999), in that the derivatives of $M(\theta)$ are estimated along with $M(\theta)$ and these estimated derivatives are exactly equal to the derivatives of the estimated $M(\theta)$. This property is referred to as self-consistency and ensures that no logical contradictions (regarding, for example, the locations of local extrema) arise from simultaneous inspection of the estimated $M(\theta)$ and the estimated derivatives.

While compound estimation is described in detail by Charnigo and Srinivasan (2011a), we present a few highlights here for convenience. Compound estimation begins by defining a grid of "centering points" $\{a_1, a_2, \ldots, a_I\}$ and using an auxiliary nonparametric regression method

to acquire estimates of $M(\theta)$ and its derivatives at the centering points. In this technical report, we shall use local regression as the auxiliary nonparametric regression method, as implemented in Loader's (1999) locfit package for R statistical software with nearest neighbor parameter 0.20, local polynomial degree 3, and rectangular kernelfunction. We shall also take $I := 27$ and $a_j := (180/27)(j - 0.5)$ for $j \in \{1, 2, \ldots, 27\}$.

Let $\widetilde{c}_{k,a_j,M}$ denote the local regression estimate of the k^{th} derivative of $M(\theta)$ evaluated at a_j, divided by $k!$, for $k \in \{0, 1, 2, 3\}$. We refer to this as a "pointwise estimate". This pointwise estimate has the linear representation $\sum_{i=1}^{G} \widetilde{L}_k(\theta_i, a_j)\{M(\theta_i) + \epsilon_i\}$ for a real-valued function $\widetilde{L}_k(x, y)$ of two real variables x and y, a fact that will be useful later. Compound estimation continues by defining an estimate of $M(\theta)$ as $\sum_{i=1}^{I} \exp[-\beta(\theta - a_i)^2]\{\sum_{k=0}^{3} \widetilde{c}_{k,a_i,M}(\theta - a_i)^k\} / \sum_{i=1}^{I} \exp[-\beta(\theta - a_i)^2]$ and an estimate of its k^{th} derivative as the k^{th} derivative of the aforementioned quantity, for $k \in \{0, 1, 2\}$. In this technical report, we shall take $\beta := 100$. Note that, even though local regression is not self-consistent, compound estimates of $M(\theta)$ and its derivatives are automatically self-consistent, by definition. This would be true, by the way, regardless of how the pointwise estimates were defined.

The principle of minimizing L^2 distance. Let $\widehat{\mu}_k(\theta)$ denote the vector-valued function whose components are the compound estimates of the k^{th} derivatives of $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$, and $M_{34}(\theta)$, for $k \in \{0, 1, 2\}$. Let $\mu_{k,p}(\theta)$ denote the vector-valued function whose components are the actual k^{th} derivatives of $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$, and $M_{34}(\theta)$, for $k \in \{0, 1, 2\}$, for nanoparticles with property p. In this technical report, we shall let $p \in \{1, 2, 3, 4\}$ correspond to agglomeration levels of 0%, 25%, 75%, and 100% respectively. Agglomeration levels refer to the percentages of single nanoparticles; the Appendix of Charnigo et al (2011) provides technical details regarding assumptions on the geometries of non-single nanoparticles. We assume that

$\mu_{k,p}(\theta)$ is known, with negligible error, based on prior experimentation on nanoparticles whose properties were known.

Consider the weighted L^2 distance $w_1 \sum_{i=1} ||\widehat{\mu}_0(\theta_i) - \mu_{0,p}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_0(\theta_i, \theta_j)^2 +$ $w_2 \sum_{i=1} ||\widehat{\mu}_1(\theta_i) - \mu_{1,p}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_1(\theta_i, \theta_j)^2 + w_3 \sum_{i=1} ||\widehat{\mu}_2(\theta_i) - \mu_{2,p}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_2(\theta_i, \theta_j)^2$, where $\widehat{L}_k(x, y)$ is a real-valued function of two real variables x and y such that $\sum_{i=1}^{G} \widehat{L}_k(\theta_i, \theta)\{M(\theta_i) + \epsilon_i\}$ is the compound estimate of the k^{th} derivative of $M(\theta)$. If the weighted L^2 distance is minimized at $p = 1$, then we infer (perhaps incorrectly) that the unknown agglomeration level is 0% for the nanoparticles to be characterized. If the weighted L^2 distance is minimized at $p = 2$, $p = 3$, or $p = 4$, then we infer (again, perhaps incorrectly) that the unknown agglomeration level is 25%, 75%, or 100% respectively for the nanoparticles to be characterized.

Note that the inference regarding the unknown agglomeration level depends on the choices of w_1, w_2, and w_3. In this technical report, we consider $(w_1, w_2, w_3) \in \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}$. The latter choice uses the estimated scattering profiles and their first two derivatives simultaneously. The denominators of $\sum_{i,j} \widehat{L}_k(\theta_i, \theta_j)^2$ in the weighted L^2 distance are proportional to the variances of the estimated scattering profiles and their first two derivatives, so that each of these objects makes a comparable contribution to the weighted L^2 distance when $(w_1, w_2, w_3) = (1, 1, 1)$.

The concept of Steinization. Charnigo and Srinivasan (2011b) provide a detailed review of Stein's phenomenon, including heuristic explanations for it and the roles of Stein's phenomenon in both the development of new methods for statistical inference and modern data analysis. We offer a few highlights here for convenience. Stein (1956) showed that, if one wished to estimate parameters $\tau_1, \tau_2, \ldots, \tau_D$ for some $D \geq 3$ upon observing normally distributed and uncorrelated $X_1, X_2, \ldots, X_D$ with means $\tau_1, \tau_2, \ldots, \tau_D$ and unit variances, there were better estimates of

$\tau_1, \tau_2, \ldots, \tau_D$ available than $X_1, X_2, \ldots, X_D$ themselves. More specifically, the expected value of the L^2 distance between $\tau_1, \tau_2, \ldots, \tau_D$ and $X_1, X_2, \ldots, X_D$ is greater than that between $\tau_1, \tau_2, \ldots, \tau_D$ and $\widehat{\tau}_1, \widehat{\tau}_2, \ldots, \widehat{\tau}_D$ with $\widehat{\tau}_j := \{1 - b/(\sum_{k=1}^{D} X_k^2 + a)\}X_j$ for $j \in \{1, 2, \ldots, D\}$ and appropriately chosen constants a and b. James and Stein (1961) proposed, more specifically, taking $b := (D-2)$ and $a := 0$. Note that $\widehat{\tau}_j$ is obtained, essentially, by nonlinearly shrinking X_j partway toward zero. Stein's phenomenon is so named because one would not anticipate that, for example, $X_2, \ldots, X_D$ should have anything to do with estimation of τ_1.

Efron and Morris (1971, 1972, 1973, 1975) subsequently published a series of papers on Stein's phenomenon and its connections to Bayesian inference. Among other contributions, they advocated nonlinearly shrinking X_j partway toward $\bar{X} := \sum_{i=1}^{D} X_k/D$, rather than toward zero, by defining $\widehat{\tau}_j := \bar{X} + \{1 - (D-3)/(\sum_{k=1}^{D}[X_k - \bar{X}]^2)\}(X_j - \bar{X})$ for $j \in \{1, 2, \ldots, D\}$. The aforementioned definitions of $\widehat{\tau}_j$ can also be adapted to situations in which $X_1, X_2, \ldots, X_D$ are correlated and/or have different from unit variances.

The theoretical justification of Stein's (1956) phenomenon uses the fact that $\tau_1, \tau_2, \ldots, \tau_D$ are the means of the observations $X_1, X_2, \ldots, X_D$. However, some nonlinear shrinkage may still be of benefit even if the means of $X_1, X_2, \ldots, X_D$ are not exactly equal to $\tau_1, \tau_2, \ldots, \tau_D$ (no "bias", in statistical parlance), provided that the means of $X_1, X_2, \ldots, X_D$ are "close" to $\tau_1, \tau_2, \ldots, \tau_D$ in relation to the variances of $X_1, X_2, \ldots, X_D$ (low bias). This idea, developed theoretically in the Ph.D. thesis of Feng (2013), is proposed for application to nanoparticle characterization in the present technical report.

Steinization of pointwise estimates across centering points. The first way in which we attempt to employ Steinization entails letting $\widetilde{c}_{k,a_1,M}, \widetilde{c}_{k,a_2,M}, \ldots, \widetilde{c}_{k,a_9,M}$ play the role of $X_1, X_2, \ldots, X_D$ with $D = 9$ for each $k \in \{0, 1, 2, 3\}$. More specifically, let $\bar{c}_{k,a_1,\ldots,a_9,M} :=$

$\sum_{m=1}^{9} \widetilde{c}_{k,a_m,M}/9$ and $\widetilde{\Sigma}_{k,a_1,\ldots,a_9,M}$ denote the estimated covariance matrix of $\widetilde{c}_{k,a_1,M}, \widetilde{c}_{k,a_2,M}, \ldots, \widetilde{c}_{k,a_9,M}$, defined in its mn element by $\sum_{i=1}^{G} \widetilde{L}_k(\theta_i, a_m) L_k(\theta_i, a_n)$ multiplied by the average of the residual mean squares in estimating $M_{11}(\theta), M_{12}(\theta), M_{33}(\theta), M_{34}(\theta)$.

Put $\widehat{c}_{k,a_j,M} := \bar{c}_{k,a_1,\ldots,a_9,M} + \{1 - 6/(\widetilde{c}_{k,a_j,M} - \bar{c}_{k,a_1,\ldots,a_9,M})' \widetilde{\Sigma}^{-1}_{k,a_1,\ldots,a_9,M} (\widetilde{c}_{k,a_j,M} - \bar{c}_{k,a_1,\ldots,a_9,M})\}^{+}$ $(\widetilde{c}_{k,a_j,M} - \bar{c}_{k,a_1,\ldots,a_9,M})$ for $k \in \{0,1,2,3\}$ and $j \in \{1,2,\ldots,9\}$, where $'$ denotes a vector transpose, $^{-1}$ denotes a matrix inverse, and $^{+}$ denotes the positive part. Thus, we nonlinearly shrink each $\widetilde{c}_{k,a_j,M}$ toward the average of $\widetilde{c}_{k,a_1,M}, \widetilde{c}_{k,a_2,M}, \ldots, \widetilde{c}_{k,a_9,M}$. This form of Steinization is illustrated in Figure 1 of the present technical report.

Also, in the event that $\widetilde{\Sigma}_{3,a_1,\ldots,a_9,M}$ is not invertible, we put $\widehat{c}_{3,a_j,M} := \bar{c}_{3,a_1,\ldots,a_9,M} + \{1 - 6/(\widetilde{c}_{2,a_j,M} - \bar{c}_{2,a_1,\ldots,a_9,M})' \widetilde{\Sigma}^{-1}_{2,a_1,\ldots,a_9,M} (\widetilde{c}_{2,a_j,M} - \bar{c}_{2,a_1,\ldots,a_9,M})\}^{+} (\widetilde{c}_{3,a_j,M} - \bar{c}_{3,a_1,\ldots,a_9,M})$. In words, the shrinkage factor for $\widehat{c}_{3,a_j,M}$ is "borrowed" from $\widehat{c}_{2,a_j,M}$ in the event that $\widetilde{\Sigma}_{3,a_1,\ldots,a_9,M}$ is not invertible.

We proceed similarly to define Steinized versions of $\widetilde{c}_{k,a_{10},M}, \widetilde{c}_{k,a_{11},M}, \ldots, \widetilde{c}_{k,a_{18},M}$ and $\widetilde{c}_{k,a_{19},M}, \widetilde{c}_{k,a_{20},M}, \ldots, \widetilde{c}_{k,a_{27},M}$. Then, instead of using $\widetilde{c}_{k,a_1,M}, \widetilde{c}_{k,a_2,M}, \ldots, \widetilde{c}_{k,a_{27},M}$ when defining compound estimates of $M(\theta)$ and its derivatives, we use $\widehat{c}_{k,a_1,M}, \widehat{c}_{k,a_2,M}, \ldots, \widehat{c}_{k,a_{27},M}$.

Let $\widehat{\mu}_{k,Stein}(\theta)$ denote the vector-valued function whose components are the compound estimates (with Steinization of the pointwise estimates) of the k^{th} derivatives of $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$, and $M_{34}(\theta)$, for $k \in \{0,1,2\}$. We can use the weighted L^2 distance $w_4 \sum_{i=1} ||\widehat{\mu}_{0,Stein}(\theta_i) - \mu_{0,p}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_0(\theta_i,\theta_j)^2 + w_5 \sum_{i=1} ||\widehat{\mu}_{1,Stein}(\theta_i) - \mu_{1,p}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_1(\theta_i,\theta_j)^2 +$ $w_6 \sum_{i=1} ||\widehat{\mu}_{2,Stein}(\theta_i) - \mu_{2,p}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_2(\theta_i,\theta_j)^2$ to make an inference about the unknown agglomeration level by determining for which $p \in \{1,2,3,4\}$ this weighted L^2 distance is minimized.

Steinization of pointwise estimates across orders of derivatives and response variables.

The second way in which we attempt to employ Steinization entails letting normalized versions of $\widetilde{c}_{0,a_j,M_{11}}, \widetilde{c}_{1,a_j,M_{11}}, \ldots, \widetilde{c}_{3,a_j,M_{34}}$ play the role of $X_1, X_2, \ldots, X_D$ with $D = 16$ for each $j \in \{1, 2, \ldots, 27\}$. The normalization entails division by $\sqrt{\sum_{i=1}^{G} \widetilde{L}_k(\theta_i, a_j)^2 \widehat{\sigma}^2}$, with $\widehat{\sigma}^2$ defined as above. Let $\widetilde{d}_{0,a_j,M_{11}}, \widetilde{d}_{1,a_j,M_{11}}, \ldots, \widetilde{d}_{3,a_j,M_{34}}$ denote the normalized versions of $\widetilde{c}_{0,a_j,M_{11}}, \widetilde{c}_{1,a_j,M_{11}}, \ldots, \widetilde{c}_{3,a_j,M_{34}}$, $\bar{d}_j$ their average, and $\widetilde{\Sigma}_j$ their estimated covariance matrix. This matrix is block diagonal, with nonzero entries of the form $\sum_{i=1}^{G} \widetilde{L}_{k_1}(\theta_i, a_j) \widetilde{L}_{k_2}(\theta_i, a_j) / \sqrt{\sum_{i=1}^{G} \widetilde{L}_{k_1}(\theta_i, a_j)^2 \sum_{i=1}^{G} \widetilde{L}_{k_2}(\theta_i, a_j)^2}$.

Put $\widehat{d}_{k,a_j,M} := \bar{d}_j + \{1 - 13/(\widetilde{d}_{k,a_j,M} - \bar{d}_j)' \widetilde{\Sigma}_j^{-1} (\widetilde{d}_{k,a_j,M} - \bar{d}_j)\}^+ \ (\widetilde{d}_{k,a_j,M} - \bar{d}_j)$ for $k \in \{0, 1, 2, 3\}$. Thus, we nonlinearly shrink each of $\widetilde{d}_{0,a_j,M_{11}}, \widetilde{d}_{1,a_j,M_{11}}, \ldots, \widetilde{d}_{3,a_j,M_{34}}$ toward their average. Then put $\widehat{c}_{k,a_j,M} := \widehat{d}_{k,a_j,M} \sqrt{\sum_{i=1}^{G} \widetilde{L}_k(\theta_i, a_j)^2 \widehat{\sigma}^2}$ and, instead of using $\widetilde{c}_{0,a_j,M_{11}}, \widetilde{c}_{1,a_j,M_{11}}, \ldots, \widetilde{c}_{3,a_j,M_{34}}$ when defining compound estimates of $M(\theta)$ and its derivatives, we use $\widehat{c}_{0,a_j,M_{11}}, \widehat{c}_{1,a_j,M_{11}}, \ldots, \widehat{c}_{3,a_j,M_{34}}$. This form of Steinization is illustrated in Figure 2 of the present technical report.

As before, we can use the weighted L^2 distance $w_4 \sum_{i=1} ||\widehat{\mu}_{0,Stein}(\theta_i) - \mu_{0,p}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_0(\theta_i, \theta_j)^2 + w_5 \sum_{i=1} ||\widehat{\mu}_{1,Stein}(\theta_i) - \mu_{1,p}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_1(\theta_i, \theta_j)^2 + w_6 \sum_{i=1} ||\widehat{\mu}_{2,Stein}(\theta_i) - \mu_{2,p}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_2(\theta_i, \theta_j)^2$ to make an inference about the unknown agglomeration level.

Numerical experimentation. Using a numerical model proposed by Aslan et al (2005) and validated by Venkata (2006), we generated $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$, and $M_{34}(\theta)$ for $\theta \in \{2, 3, \ldots, 178\}$ corresponding to nanoparticles with diameters of 50 nanometers (nm) and agglomeration levels of 0%, 25%, 75%, and 100%. We assume, as in Charnigo et al (2011), that the radiation wavelength is 514.5 nm and has a 23-degree angle of incidence, the prism and substrate are made of sapphire with refractive index 1.77304, the substrate is coated with a 20 nm gold film having (complex) refractive index $0.50 + 1.86i$, and the scatterers are gold

spherical nanoparticles having refractive index $0.50 + 1.86i$. Agglomeration levels refer to the percentages of single nanoparticles; the Appendix of Charnigo et al (2011) provides technical details regarding assumptions on the geometries of non-single nanoparticles.

We applied compound estimation to $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$, and $M_{34}(\theta)$ for $\theta \in \{2, 3, \ldots, 178\}$ in order to obtain estimates of $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$, and $M_{34}(\theta)$ and their first two derivatives for $\theta \in [0, 180]$. These estimates were used to define $\mu_{0,p}(\theta)$, $\mu_{1,p}(\theta)$, and $\mu_{2,p}(\theta)$ for the aforementioned weighted L^2 distances.

We then generated alternate versions of $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$, and $M_{34}(\theta)$ for $\theta \in \{2, 3, \ldots, 178\}$ based on 14 prescribed deviations from the numerical model referred to by Charnigo et al (2011) as "disturbances" and specifically enumerated in Section 4.1 of their manuscript. Both the original and alternate versions of $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$, and $M_{34}(\theta)$ for $\theta \in \{2, 3, \ldots, 178\}$ were subjected to 10 different realizations of stochastic noise, to obtain a total of 600 data sets: 150 data sets in which the true agglomeration level was 0%, 150 with 25% agglomeration, 150 with 75%, and 150 with 100%. The stochastic noise added to $M(\theta_1), \ldots, M(\theta_G)$ in any particular data set was drawn from a normal distribution with mean zero and standard deviation 0.50 times $\sqrt{\{\sum_{i=1}^{G} M(\theta_i)^2 - [\sum_{i=1}^{G} M(\theta_i)]^2/G\}/(G-1)}$. Note that the 0.50 is a *much* higher noise level than the 0.03 considered by Charnigo et al (2007, 2011). Thus, the present investigation considers a scenario in which Steinization is potentially helpful, inasmuch as estimation bias is very low relative to estimation variance, since the latter is quite high at a 0.50 noise level. This noise level is illustrated in Figure 3 of the present technical report.

We next applied compound estimation without Steinization to each of the 600 data sets and used the principle of minimizing weighted L^2 distance to classify each of the 600 data

sets. This was done for $(w_1, w_2, w_3) \in \{(1,0,0),(0,1,0),(0,0,1),(1,1,1)\}$, as described above. We also applied compound estimation with Steinization across centering points to each of the 600 data sets and again employed the principle of minimizing weighted L^2 distance to classify each of the 600 data sets. This was done for $(w_4, w_5, w_6) \in \{(1,0,0),(0,1,0),(0,0,1),(1,1,1)\}$, as described above. Because this was a numerical experiment, the "correct" classification was known to us, and so we could ascertain correct classification rates with and without Steinization for each choice of weights.

These findings are reported in Table 1, which we describe in the Results section. We also mention here that we calculated not only correct classification rates but also the average weighted L^2 distance from $\widehat{\mu}_0(\theta), \widehat{\mu}_1(\theta), \widehat{\mu}_2(\theta)$ or $\widehat{\mu}_{0,Stein}(\theta), \widehat{\mu}_{1,Stein}(\theta), \widehat{\mu}_{2,Stein}(\theta)$ to $\mu_{0,correct}(\theta), \mu_{1,correct}(\theta), \mu_{2,correct}(\theta)$ and $\mu_{0,incorrect}(\theta), \mu_{1,incorrect}(\theta), \mu_{2,incorrect}(\theta)$. The "correct" subscript is $p = 1$ for the first 150 data sets, $p = 2$ for the next 150, $p = 3$ for the third 150, and $p = 4$ for the last 150. The "incorrect" subscripts are $p \in \{2,3,4\}$ for the first 150 data sets, $p \in \{1,3,4\}$ for the next 150, $p \in \{1,2,4\}$ for the third 150, and $p \in \{1,2,3\}$ for the last 150.

The preceding numerical experiment was also repeated with a 0.25 noise level, which is still rather high, and with Steinization across orders of derivatives and response variables rather than across centering points. These findings are reported in Tables 2, 3, and 4, which we describe in the Results section.

3 Results

Steinization of pointwise estimates across centering points. Table 1 displays findings at the 0.50

noise level when pointwise estimators are Steinized across clusters of nine centering points for each outcome and each order of derivative. The average L^2 distance from scattering profiles estimated at the 0.50 noise level to the actual scattering profiles corresponding to correct classifications was reduced by 7% with Steinization. Yet, Steinization did not increase the frequency with which minimizing the L^2 distance over different nanoparticle configurations yields a correct classification, which remained around 90%. This may have been because Steinization also reduced by 3% the average L^2 distance from estimated scattering profiles to the actual scattering profiles corresponding to incorrect classifications. There were only six data sets out of the 600 for which the classification was modified by Steinization. In one instance, Steinization changed an incorrect classification to a correct one; in five instances, Steinization changed a correct classification to an incorrect one.

When first derivatives are considered instead of undifferentiated scattering profiles, Steinization reduced by 25% the average L^2 distance from the estimated derivatives to the actual derivatives corresponding to correct classifications, versus 20% for the actual derivatives for incorrect classifications. Steinization slightly increased the correct classification rate, from about 73% to 75%. There were 61 data sets for which the classification was altered by Steinization, 34 in which Steinization changed an incorrect classification to correct, 26 in which Steinization denatured a correct classification, and one in which one incorrect classification converted to another incorrect classification. Interestingly, all 15 of the 61 with a true aggregation level of 0% were misclassified without but correctly classified with Steinization. Among 21 with a true aggregation level of 25%, eight instances saw Steinization change an incorrect classification to a correct one and 13 saw Steinization change a correct classification to an incorrect one. However, in all 21 instances, Steinization changed the classification to one with a lower

agglomeration level.

As for second derivatives, Steinization diminished by 31% the average L^2 distance from the estimated derivatives to the actual derivatives corresponding to correct classifications, rather than 30% for the actual derivatives for incorrect classifications. Steinization slightly increased the correct classification rate, from about 44% to 46%. There were 78 data sets for which the classification was modified by Steinization, 31 in which Steinization changed an incorrect classification to correct, 18 in which Steinization denatured a correct classification, and 29 in which one incorrect classification converted to another incorrect classification. In all but two instances, Steinization changed the classification to one with a lower agglomeration level.

When scattering profiles, first derivatives, and second derivatives are considered simultaneously, Steinization reduced by 19% the average weighted L^2 distance from the estimated objects to the actual objects corresponding to correct classifications, versus 10% for the actual objects for incorrect classifications. Yet, Steinization did not increase the frequency with which minimizing the L^2 distance over different nanoparticle configurations yields a correct classification, which remained around 91%. There were 12 data sets for which the classification was altered by Steinization, four in which Steinization changed an incorrect classification to a correct one and eight in which Steinization changed a correct classification into an incorrect one.

Table 2 shows findings at the 0.25 noise level. As anticipated, average L^2 distances are uniformly and considerably lower than at the 0.50 noise level. However, Steinization still decreases these distances, both when the actual objects correspond to correct classifications and when they do not. While correct classification rates are uniformly higher than at the 0.50 noise level (and considerably so for the first and second derivatives), Steinization has little impact

on them. There were no data sets in which classification based on undifferentiated scattering profiles was affected by Steinization, 14 in which classification based on first derivatives was altered, 35 in which classification based on second derivatives was modified, and two in which Steinization impacted classification from the scattering profiles and their first two derivartives simultaneously.

Steinization of pointwise estimates across orders of derivatives and response variables. Table 3 displays findings at the 0.50 noise level when pointwise estimators are Steinized across each outcome and each order of derivative for each centering point. Steinization reduces average L^2 distances by 1% to 2%, but correct classification rates are virtually unaltered. There were 11 data sets in which classification based on undifferentiated scattering prorfiles was affected by Steinization, four in which classification based on first derivatives was altered, nine in which classification based on second derivatives was modified, and eight in which Steinization impacted classification from the scattering profiles and their first two derivatives simultaneously. Among these 32 data sets, there were 25 in which Steinization changed a classification of 100% agglomeration to 75% agglomeration.

Table 4 shows findings at the 0.25 noise level. As anticipated, average L^2 distances are uniformly and considerably lower than at the 0.50 noise level. However, Steiniation has negligible impact on these distances, regardless of whether the actual objects correspond to correct classifications. Although correct classification rates are uniformly higher than at the 0.50 noise level (and considerably so for the first and second derivatives), Steinization has little effect on them. There were four data sets in which classification based on undifferentiated scattering prorfiles was affected by Steinization, one in which classification based on first derivatives was altered, three in which classification based on second derivatives was modified, and none

in which Steinization impacted classification from the scattering profiles and their first two derivatives simultaneously.

4 Discussion

The findings from our numerical experimentation were decidedly mixed. Although Steinization across centering points substantially reduced error in estimating the derivatives of scattering profiles, there was not a commensurate improvement in the correct classification rates from using the estimated derivatives with Steinization versus from using the estimated derivatives without Steinization. Moreover, the slight uptick from 263/600 correct classifications to 276/600 based on second derivatives, while ostensibly mildly encouraging, is of limited practical import because correct classification rates considerably higher than 276/600 are available from first derivatives and scattering profiles themselves without recourse to Steinization. There is some possibility that Steinization might confer greater benefit if the noise level were higher than 0.50, but this seems more of a theoretical than a practical consideration. As shown in Figure 3, a noise level of 0.50 is already very high. One has difficulty imagining that an experimenter would regard as viable an approach to nanoparticle characterization based on scattering profiles, much less their derivatives, if the data were of worse quality (i.e., more noisy and/or disturbed) than shown in Figure 3.

We reiterate that the noise levels of 0.25 and 0.50 considered herein are *much* higher than the noise level of 0.03 considered by Charnigo et al (2007, 2011). As such, we are not surprised that derivatives of scattering profiles did not yield more accurate nanoparticle characterization than the scattering profiles themselves in the present numerical experimentation (although a

slight advantage was conferred by looking at scattering profiles and their derivatives simultaneously). This conclusion appears to differ from the findings of Charnigo et al (2007, 2011), but estimation of a derivative is impaired much more than estimation of the underlying curve in the presence of high noise. Indeed, if the noise level had been as low as 0.03, the correct classification rates would have been near perfect for all orders of derivative considered herin, with or without Steinization. Incidentally, such near-perfect correct classification rates do not conflict with the findings of Charnigo et al (2007, 2011), which were based mostly on separate analyses of $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$, and $M_{34}(\theta)$ with their respective derivatives. Although suppressed from Tables 1 through 4 for brevity, results based on L^2 distances involving $M_{11}(\theta)$ and its respective derivatives only (rather than all four scattering profiles and their respective deriatives) were considerably worse than results based on L^2 distances involving $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$, and $M_{34}(\theta)$ and their respective derivatives simultaneously. The same is true for L^2 distances involving $M_{12}(\theta)$ and its respective derivatives only and, to a lesser degree, $M_{33}(\theta)$ only and $M_{34}(\theta)$ only. In sum, the results described in this technical report are not at variance with but are complementary to the results of Charnigo et al (2007, 2011).

In contrast to Steinization across centering points, Steinization across orders of derivatives and response variables had almost no detectible effect either on the quality of estimation or on the correct classification rates. This is apparently because the shrinkage factor for this type of Steinization was much closer to 1 (i.e., almost as if there were no Steinization at all). One may wonder whether a more sophisticated calculation of the shrinkage factor, taking into account the bias inherent to estimation in the nonparametric regression setting, would be of consequence. Indeed, a more sophisticated calculation is possible, and Feng (2013) theoretically examined this possibility in her Ph.D. thesis. However, we do not anticipate that

a more sophisticated calculation would be of consequence here because its effect would be to bring the shrinkage factor even closer to 1. By the same token, one may wonder whether a more sophisticated calculation would be of consequence with Steinization across centering points. That answer is less obvious. Clearly, Steinization across centering points had some benefit in reducing the error in estimating the derivatives of the scattering profiles; a more sophisticated calculation could potentially reduce the error further, to the point that correct classification results improved as a consequence. This possibility bears closer examination in ongoing research.

There are, of course, other ways to exploit Stein's phenomenon besides the two considered herein. For instance, one may Steinize across orders of derivatives and centering points simultaneously, or across response variables and centering points simultaneously. Perhaps a more creative idea for ongoing research is to perform Steinization across multiple data sets presumed to arise from a common collection of nanoparticles (or, at worst, multiple collections of nanoparticles with common properties). To illustrate this idea, consider again Figure 2. However, this time imagine that the black and red curves correspond not to estimates of $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$, and $M_{34}(\theta)$ at centering points from a single data set but rather estimates of $M_{11}(\theta)$ at centering points from different data sets. If these data sets arise from a common collection of nanoparticles, then the black and red curves are actually estimating the same object, so there is a very compelling motivation to adjust the black curve toward the red curves. Indeed, even if one were unaware of Stein's phenomenon, the idea of somehow averaging the black curve and the red curves would be appealing inasmuch as averaging tends to reduce error when the quantities involved are estimating the same object.

The preceding point, however, is also the crux of a major conceptual question: If the black

and red curves in Figure 2 really are estimating the same object, then why Steinize across multiple data sets instead of simply averaging ? If one does Steinize, then one adjusts the black curve and each of the red curves so that they are all mutually closer. Yet, one still has four distinct curves and will subsequently obtain four distinct estimates of $M_{11}(\theta)$ over the continuum of observation angles. Each one of these four estimates of $M_{11}(\theta)$ can be used to classify the nanoparticles, but there is no guarantee that these four classifications will be the same. Quite possibly, for example, three classifications may be correct and one incorrect. Now contrast the preceding approach with taking a simple average of the black curve and the red curves to get a single curve and a single estimate of $M_{11}(\theta)$ over the continuum of observation angles. This single estimate of $M_{11}(\theta)$ can be used to classify the nanoparticles, and in terms of the original data sets the result will either be four correct classifications out of four or none. So, unless one anticipates a fairly sizable probability of incorrect classification through simple averaging of the black curve and the red curves, the expected correct classification rate from simple averaging may be superior to that from Steinization across multiple data sets.

Notwithstanding this conceptual question, there remain multiple avenues for future research that may be profitably pursued. One of them is, indeed, to perform the simple averaging over multiple data sets but then Steinize the simple averages themselves, either across centering points or across response variables and orders of derivatives. Another is to make the classifications probabilistic. That is, instead of merely identifying the most plausible configuration for the nanoparticles in question, assign indices of plausibility to the various possible configurations. If this is done, then some further benefits of Steinization may manifest. For instance, even if the correct classification rate is unchanged from 559/600 after Steinization is performed, perhaps the experimenter may go from being "80% confident" in the classifications

(on average) to "90% confident". Of course, phrases such as "80% confident" and "90% confident" must be formalized, but the point is that probabilistic classifications may become more definitive after Steinization. And, presumably, a greater level of confidence is desirable if such classifications will eventually be relied upon to engineer nanostructured surfaces for efficient nano-TPV power generation.

Acknowledgments

This material is based upon work supported by the Army Research Office under Grant No. W911NF-12-1-0422. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Army Research Office.

References

[1] Aslan, M. M., Mengüç, M. P., and Videen G. Characterization of metallic nanoparticles via surface wave scattering: B. Physical concept and numerical experiments (2005). *Journal of Quantitative Spectroscopy and Radiative Transfer*, **93**, 207–217.

[2] Baxter J. et al. (2009). Nanoscale design to enable the revolution in renewable energy. *Energy and Environmental Science*, **2**, 559-588.

[3] Bergman T.L., Lavine A.S., Incropera F.P., and DeWitt D.P. (2011). *Fundamentals of Heat and Mass Transfer*, 7^{th} edition. Wiley, New York.

[4] Bohren, C. and Huffman, D. (1998). *Absorption and Scattering of Light by Small Particles*. Wiley, New York.

[5] Charnigo, R., Francoeur, M., Kenkel, P., Menguc, M.P., Hall, B., and Srinivasan, C. (2012). Credible intervals for nanoparticle characteristics. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **113**: 182–193.

[6] Charnigo, R., Francoeur, M., Kenkel, P., Menguc, M.P., Hall, B., and Srinivasan, C. (2011). Estimating quantitative features of nanoparticles using multiple derivatives of scattering profiles. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **112**: 1369–1382.

[7] Charnigo, R., Francoeur, M., Mengüç, M. P., Brock, A., Leichter, M., and Srinivasan, C. (2007). Derivatives of scattering profiles: tools for nanoparticle characterization. *Journal of the Optical Society of America A*, **24**: 2578–2589.

[8] Charnigo, R. and Srinivasan, C. (2011a). Self-consistent estimation of mean response functions and their derivatives. *Canadian Journal of Statistics*, **39**: 280–299.

[9] Charnigo, R. and Srinivasan, C. (2011b). Stein's phenomenon. *Philosophy of Statistics*, edited by Prasanta S. Bandyopadhyay and Malcolm Forster. Elsevier, Oxford.

[10] Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**: 829–836.

[11] Efron, B. and Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators — part I: the Bayes case. *Journal of the American Statistical Association*, **66**, 807-815.

[12] Efron, B. and Morris, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators — part II: the empirical Bayes case. *Journal of the American Statistical Association*, **67**, 130-139.

[13] Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors — an empirical Bayes approach. *Journal of the American Statistical Association*, **68**, 117-130.

[14] Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its competitors. *Journal of the American Statistical Association*, **70**, 311-319.

[15] Feng, L. (2013). James-Stein type compound estimation of multiple mean response functions and their derivatives. Ph.D. thesis, University of Kentucky.

[16] Francoeur M. (2010). Near-Field Radiative Transfer: Thermal Radiation, Thermophotovoltaic Power Generation and Optical Characterization. Ph.D. thesis, University of Kentucky.

[17] Francoeur M., Mengüç M.P. and Vaillon R. (2010). Spectral tuning of near-field radiative heat flux between two thin silicon carbide films. *Journal of Physics D: Applied Physics*, **43**, 075501.

[18] Francoeur M., Vaillon R. and Mengüç M.P. (2011). Thermal impacts on the performance of nanoscale-gap thermophotovoltaic power generators. *IEEE Transactions on Energy Conversion*, **26**, 686-698.

[19] Francoeur M., Venkata P.G. and Mengüç M.P. (2007). Sensitivity analysis for characterization of gold nanoparticles and agglomerates via surface plasmon scattering patterns. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **106**, 44-55.

[20] Garcia-Etxarri A., Gomez-Medina R., Froufe-Perez L.S., Lopez C., Chantada L., Scheffold F., Aizpurua J., Nieto-Vesperinas M., and Saenz J.J. (2011). Strong magnetic response of submicron silicon particles in the infrared. *Optics Express*, **19**, 4815-4826.

[21] James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (J. Neyman, ed.), **1**, 361-379. University of California Press.

[22] Joulain K., Mulet J.-P., Marquier F., Carminati R. and Greffet J.-J. (2005). Surface electromagnetic waves thermally excited: Radiative heat transfer, coherence properties and Casimir forces revisited in the near field. *Surface Science Reports*, **57**, 59-112.

[23] Laroche M., Carminati R. and Greffet J.-J. (2006). Near-field thermophotovoltaic energy conversion. *Journal of Applied Physics*, **100**, 063704.

[24] Loader, C. (1999). *Local Regression and Likelihood*. Springer, New York.

[25] Loke V.L.Y. and Mengüç M.P. (2010). Surface waves and atomic force microscope probe-particle near-field coupling: discrete dipole approximation with surface interaction. *Journal of the Optical Society of America A*, **27**, 2293-2203.

[26] Loke V.L.Y., Mengüç M.P., and Nieminen T.A. (2011). Discrete-dipole approximation with surface interaction: Computational toolbox for MATLAB. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **112**, 1711-1725.

[27] Mengüç M.P. and Manickavasagam S. (1998). Characterization of size and structure of agglomerates and inhomogeneous particles via polarized light. *International Journal of Engineering Science*, **36**, 1569-1593.

[28] Mulet J.-P., Joulain K., Carminati R. and Greffet J.-J. (2002). Enhanced radiative heat transfer at nanometric distances. *Nanoscale and Microscale Thermophysical Engineering*, **6**, 209-222.

[29] Narayanaswamy A. and Chen G. (2003). Surface modes for near field thermophotovoltaics. *Applied Physics Letters*, **82**, 3544-3546.

[30] Novotny L. and Hecht B. (2006). *Principles of Nano-Optics*. Cambridge University Press, New York.

[31] Park K., Basu S., King W.P. and Zhang Z.M. (2008). Performance analysis of near-field thermophotovoltaic devices considering absorption distribution, *Journal of Quantitative Spectroscopy and Radiative Transfer*, **109**, 305-316.

[32] Petersen S.J., Basu S., Raeymaekers B., and Francoeur M. (2013). Tuning near-field thermal radiative properties by quantifying sensitivity of Mie resonance-based metamaterials design parameters. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **129**, 277-286.

[33] Short M.S. (2013). The Discrete Dipole Approximation with Surface Interaction for Evanescent Wave-Based Characterization of Nanostructures on a Surface with Validation Against Experimental Results. M.S. thesis, University of Utah.

[34] Short M.R., Geffrin J.-M., Vaillon R., Tortel H., Lacroix B., and Francoeur M. (2013). Evanescent wave scattering by particles on a surface: Validation of the discrete dipole approximation with surface interaction against microwave analog experiments. Submitted for publication.

[35] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (J. Neyman, ed.), **1**, 197-206. University of California Press.

[36] Venkata, P. G. (2006). Surface wave scattering from metallic nano particles and agglomerates: theoretical framework and numerical analysis. Master's thesis, University of Kentucky.

[37] Venkata, P. G., Aslan, M. M., Mengüç, M. P., and Videen, G. Surface plasmon scattering by gold nanoparticles and two-dimensional agglomerates (2007). *ASME Journal of Heat Transfer*, **129**, 60–70.

[38] Videen, G., Aslan, M. M., and Mengüç, M. P. (2005). Characterization of metallic nanoparticles via surface wave scattering: A. Theoretical framework. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **93**, 195–206.

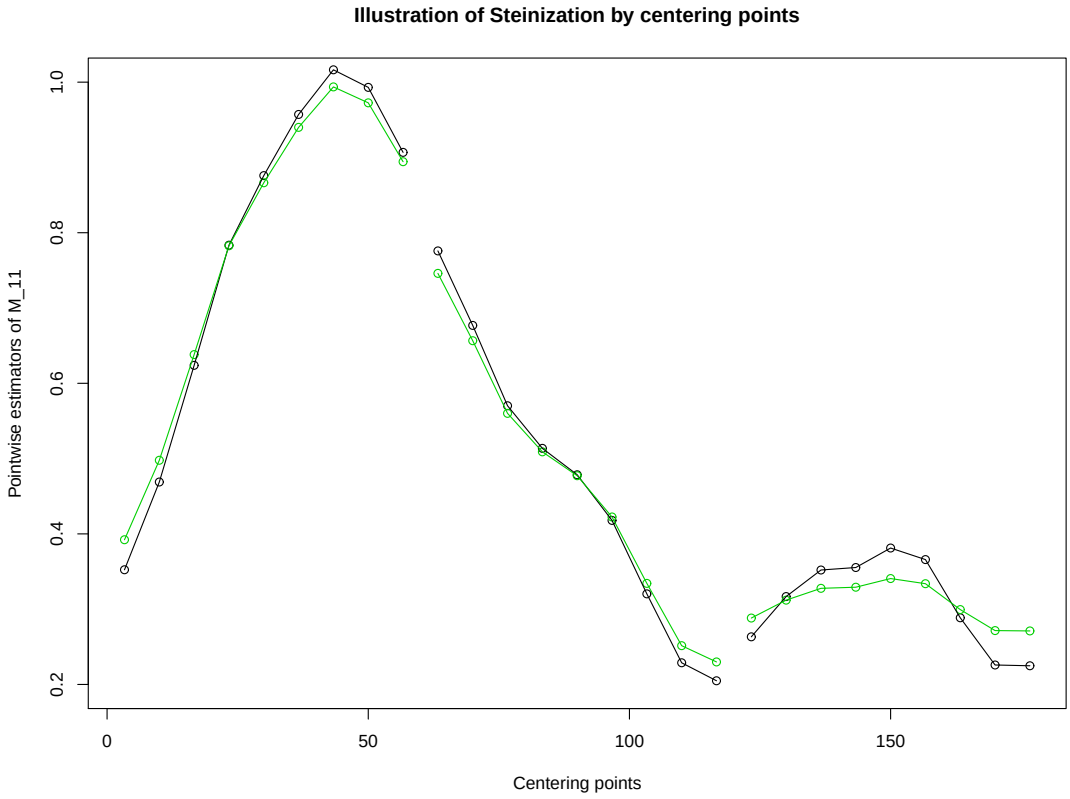
[39] West P.R., Ishii S., Naik G.V., Emani N.K., Shalaev V.M. and Boltasseva A. (2010). Searching for better plasmonic materials. *Laser and Photonics Review*, **4**, 795-808.

[40] Whale M.D. (1997). A fluctuational electrodynamics analysis of microscale radiative heat transfer and the design of microscale thermophotovoltaic devices. Ph.D. Thesis, Massachusetts Institute of Technology.

[41] Whale M.D. and Cravalho E.G. (2002). Modeling and performance of microscale thermophotovoltaic energy conversion devices. *IEEE Transactions on Energy Conversion*, **17**, 130-142.

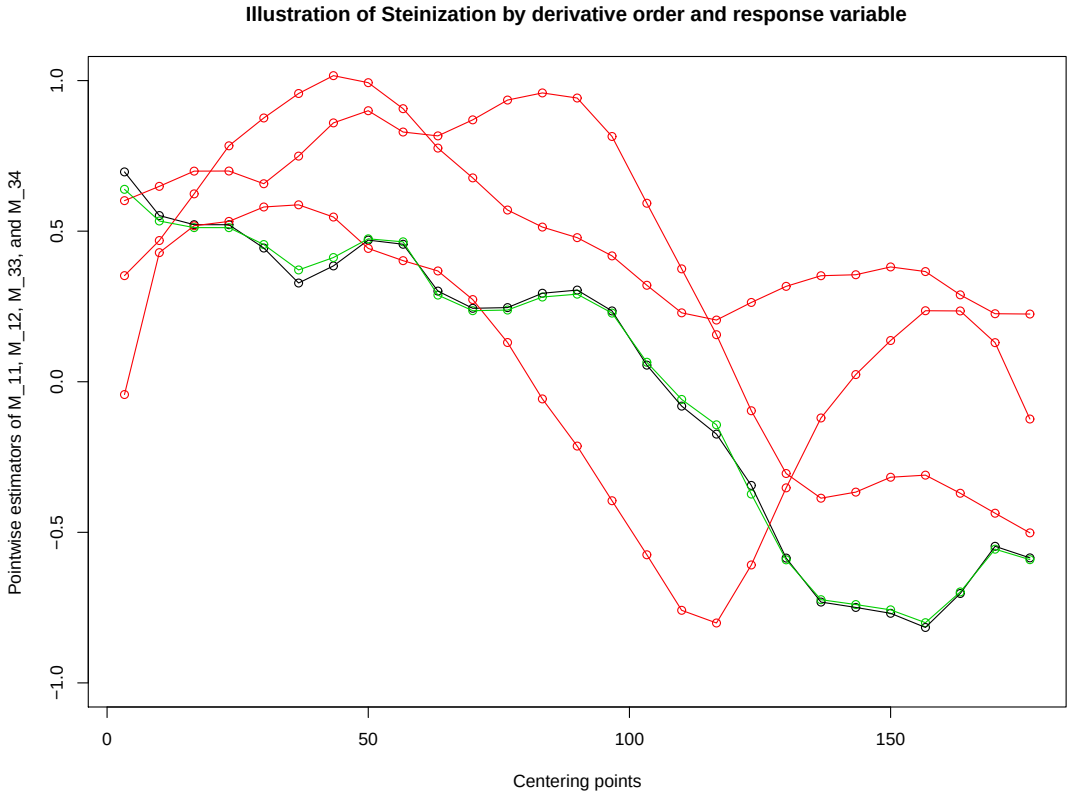
[42] Wheeler M.S., Aitchison J.S., and Mojahedi M. (2005). Three-dimensional array of dielectric spheres with isotropic negative permeability at infrared frequencies. *Physical Review B*, **72**, 193103.

Figure 1



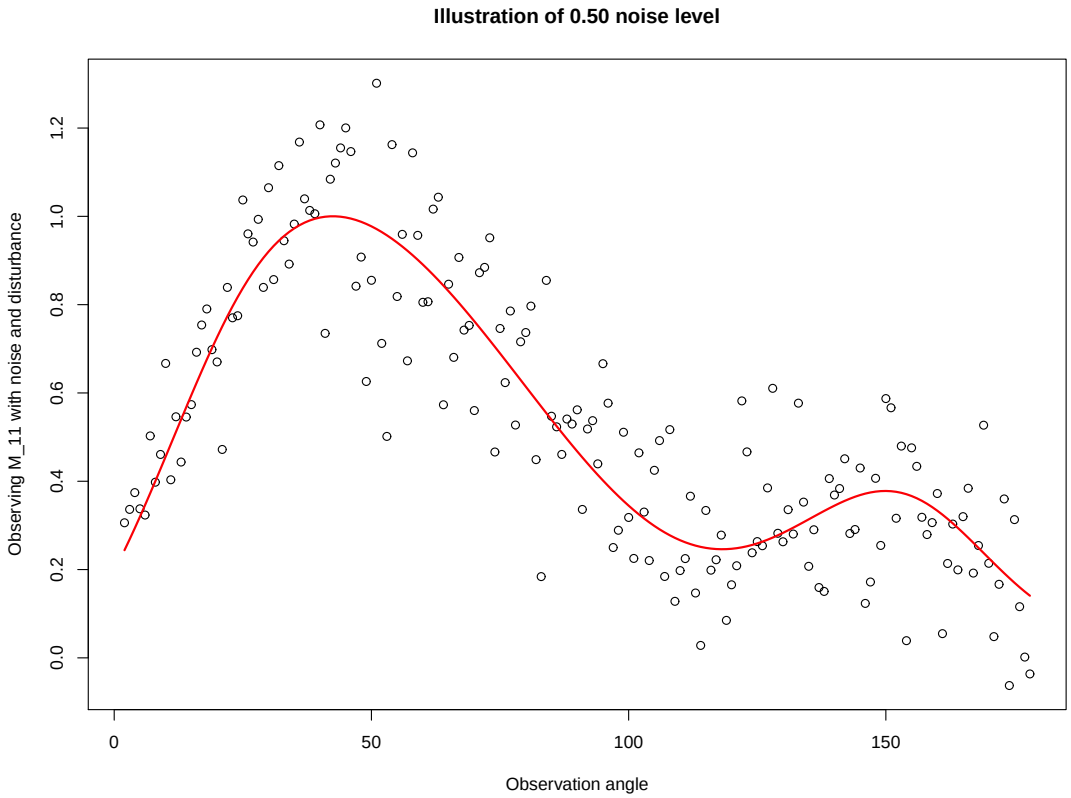
Shown in black are pointwise estimates of $M_{11}(a_1), M_{11}(a_2), \ldots, M_{11}(a_9)$ without Steinization and in green are pointwise estimates with Steinization, where $a_1, a_2, \ldots, a_9$ denote the first nine centering points for compound estimation. Note that the latter pointwise estimates represent a compromise of the former pointwise estimates with their average, so that Steinization pulls the smallest pointwise estimates upward and the largest pointwise estimates downward. Line segments connecting the pointwise estimates are included solely for ease of visualization. Likewise, pointwise estimates of $M_{11}(a_{10}), M_{11}(a_{11}), \ldots, M_{11}(a_{18})$ and $M_{11}(a_{19}), M_{11}(a_{20}), \ldots, M_{11}(a_{27})$ are shown without (black) and with (green) Steinization.

Figure 2



Shown in black are pointwise estimates of $M_{34}(a_1), M_{34}(a_2), \ldots, M_{34}(a_{27})$ without Steinization and in green are pointwise estimates with Steinization, where $a_1, a_2, \ldots, a_{27}$ denote the 27 centering points for compound estimation. Also, shown in red are pointwise estimates of $M_{11}(a_1), M_{11}(a_2), \ldots, M_{11}(a_{27})$, $M_{12}(a_1), M_{12}(a_2), \ldots, M_{12}(a_{27})$, and $M_{33}(a_1), M_{33}(a_2), \ldots, M_{33}(a_{27})$ without Steinization. Note that the pointwise estimates in green represent a compromise of the pointwise estimates in black with the pointwise estimates in red (along with 12 other sets of pointwise estimates that are not shown because their inclusion would make the graph illegible); this is most obvious at left. Line segments connecting the pointwise estimates are included solely for ease of visualization.

Figure 3



Displayed are $M_{11}(\theta_1) + \epsilon_1, M_{11}(\theta_2) + \epsilon_2, \ldots, M_{11}(\theta_G) + \epsilon_G$, where $\epsilon_1, \epsilon_2, \ldots, \epsilon_G$ embody systematic errors (from "disturbances", in the parlance of Charnigo et al, 2011) as well as stochastic errors at noise level 0.50, as defined in the Methods section. Also displayed is the curve for $M_{11}(\theta)$, without noise or disturbance.

Table 1: Numerical study results — Scenario 1

$(w_1, w_2, w_3, w_4, w_5, w_6)$	Distance to correct	Distance to incorrect	Correct classification
$(1, 0, 0, 0, 0, 0)$	0.273	0.928	541/600
$(0, 0, 0, 1, 0, 0)$	0.255	0.903	537/600
$(0, 1, 0, 0, 0, 0)$	0.202	0.255	440/600
$(0, 0, 0, 0, 1, 0)$	0.151	0.203	448/600
$(0, 0, 1, 0, 0, 0)$	0.196	0.205	263/600
$(0, 0, 0, 0, 0, 1)$	0.135	0.143	276/600
$(1, 1, 1, 0, 0, 0)$	0.671	1.389	547/600
$(0, 0, 0, 1, 1, 1)$	0.541	1.249	543/600

Scenario 1 indicates Steinization across centering points with a noise level of 0.50, as described in the Methods section. The first column presents the average, over 600 data sets, of the weighted L^2 distance $w_1 \sum_i ||\widehat{\mu}_0(\theta_i) - \mu_{0,correct}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_0(\theta_i, \theta_j)^2 + w_2 \sum_i ||\widehat{\mu}_1(\theta_i) - \mu_{1,correct}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_1(\theta_i, \theta_j)^2 + w_3 \sum_i ||\widehat{\mu}_2(\theta_i) - \mu_{2,correct}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_2(\theta_i, \theta_j)^2 + w_4 \sum_i ||\widehat{\mu}_{0,Stein}(\theta_i) - \mu_{0,correct}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_0(\theta_i, \theta_j)^2 + w_5 \sum_i ||\widehat{\mu}_{1,Stein}(\theta_i) - \mu_{1,correct}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_1(\theta_i, \theta_j)^2 + w_6 \sum_i ||\widehat{\mu}_{2,Stein}(\theta_i) - \mu_{2,correct}(\theta_i)|\,|^2 / \sum_{i,j} \widehat{L}_2(\theta_i, \theta_j)^2$, with notation as in the Methods section. The second column presents the average, over 600 data sets and 3 incorrect classifications, of the analogous quantity with $\mu_{0,correct}$, $\mu_{1,correct}$, and $\mu_{2,correct}$ replaced by $\mu_{0,incorrect}$, $\mu_{1,incorrect}$, and $\mu_{2,incorrect}$. The third column indicates the number of the 600 data sets for which minimization of weighted L^2 distance yields a correct classification.

Table 2: Numerical study results — Scenario 2

$(w_1, w_2, w_3, w_4, w_5, w_6)$	Distance to correct	Distance to incorrect	Correct classification
$(1, 0, 0, 0, 0, 0)$	0.126	0.787	552/600
$(0, 0, 0, 1, 0, 0)$	0.125	0.784	552/600
$(0, 1, 0, 0, 0, 0)$	0.056	0.110	550/600
$(0, 0, 0, 0, 1, 0)$	0.049	0.102	550/600
$(0, 0, 1, 0, 0, 0)$	0.051	0.059	374/600
$(0, 0, 0, 0, 0, 1)$	0.042	0.050	379/600
$(1, 1, 1, 0, 0, 0)$	0.233	0.957	563/600
$(0, 0, 0, 1, 1, 1)$	0.216	0.937	561/600

Scenario 2 indicates Steinization across centering points with a noise level of 0.25, as described in the Methods section. The columns are defined as in Table 1.

Table 3: Numerical study results — Scenario 3

$(w_1, w_2, w_3, w_4, w_5, w_6)$	Distance to correct	Distance to incorrect	Correct classification
$(1, 0, 0, 0, 0, 0)$	0.272	0.938	546/600
$(0, 0, 0, 1, 0, 0)$	0.269	0.927	545/600
$(0, 1, 0, 0, 0, 0)$	0.205	0.260	453/600
$(0, 0, 0, 0, 1, 0)$	0.200	0.255	453/600
$(0, 0, 1, 0, 0, 0)$	0.200	0.209	253/600
$(0, 0, 0, 0, 0, 1)$	0.196	0.204	253/600
$(1, 1, 1, 0, 0, 0)$	0.677	1.407	551/600
$(0, 0, 0, 1, 1, 1)$	0.665	1.386	551/600

Scenario 3 indicates Steinization across outcomes and orders with a noise level of 0.50, as described in the Methods section. The columns are defined as in Table 1.

Table 4: Numerical study results — Scenario 4

$(w_1, w_2, w_3, w_4, w_5, w_6)$	Distance to correct	Distance to incorrect	Correct classification
$(1, 0, 0, 0, 0, 0)$	0.125	0.784	554/600
$(0, 0, 0, 1, 0, 0)$	0.125	0.782	552/600
$(0, 1, 0, 0, 0, 0)$	0.055	0.109	543/600
$(0, 0, 0, 0, 1, 0)$	0.055	0.108	544/600
$(0, 0, 1, 0, 0, 0)$	0.050	0.058	354/600
$(0, 0, 0, 0, 0, 1)$	0.050	0.058	352/600
$(1, 1, 1, 0, 0, 0)$	0.230	0.951	559/600
$(0, 0, 0, 1, 1, 1)$	0.229	0.948	559/600

Scenario 4 indicates Steinization across outcomes and orders with a noise level of 0.25, as described in the Methods section. The columns are defined as in Table 1.